

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**



PCT

特許協力条約に基づいて公開された国際出願

<p>(51) 国際特許分類6 C12N 15/10</p>	<p>A1</p>	<p>(11) 国際公開番号 WO99/20750</p> <p>(43) 国際公開日 1999年4月29日(29.04.99)</p>
<p>(21) 国際出願番号 PCT/JP98/04772</p> <p>(22) 国際出願日 1998年10月21日(21.10.98)</p> <p>(30) 優先権データ 特願平9/289982 1997年10月22日(22.10.97) JP</p> <p>(71) 出願人 (米国を除くすべての指定国について) 株式会社 ヘリックス研究所 (HELIX RESEARCH INSTITUTE)[JP/JP] 〒292-0812 千葉県木更津市矢那1532番地3 Chiba, (JP)</p> <p>(72) 発明者 ; および</p> <p>(75) 発明者 / 出願人 (米国についてのみ) 太田紀夫(OTA, Toshio)[JP/JP] 〒292-0801 千葉県木更津市請西2-16-13-401 Chiba, (JP) 西川哲夫(NISHIKAWA, Tetsuo)[JP/JP] 〒292-0833 千葉県木更津市貝渕3-9-17-509 Chiba, (JP) サラモフ アサフ(SALAMOV, Asaf)[GB/GB] シービー10 2エイピー エセックス、サフロン ウォールデン ハーヴェイ ウェイ36 Essex, (GB) 磯貝隆夫(ISOGAI, Takao)[JP/JP] 〒292-0833 千葉県木更津市貝渕3-9-17-606 Chiba, (JP)</p>		<p>(74) 代理人 弁理士 清水初志, 外(SHIMIZU, Hatsushi et al.) 〒300-0847 茨城県土浦市卸町1-1-1 関鉄つくばビル6階 Ibaraki, (JP)</p> <p>(81) 指定国 AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO特許 (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), ユーロパ特許 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), 欧州特許 (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI特許 (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</p> <p>添付公開書類 国際調査報告書</p>
<p>(54)Title: METHOD FOR SCREENING FULL-LENGTH cDNA CLONES</p> <p>(54)発明の名称 完全長cDNAクローンの選択方法</p> <p>(57) Abstract A method for efficiently screening full-length cDNA clones which comprises: determining the base sequence in the 5'-region of each clone contained in a cDNA library prepared by a method for constructing a cDNA library involving full-length ones at a high ratio; examining the presence/absence of initiation ATG in this 5'-region and the location thereof by using an originally developed software for anticipating initiation codons in cDNA; thus exactly judging the presence/absence of the initiation codon and the location thereof; and screening the cDNAs thus judged as carrying the initiation codon from the cDNA library. Moreover, a cDNA library containing full-length ones at an extremely high ratio can be constructed by mixing the clones thus selected above.</p>		

完全長率の高いcDNAライブラリー作成法で作製したcDNAライブラリーに含まれるクローンの5'領域の塩基配列を決定し、独自に開発したcDNAの翻訳開始コドン予測ソフトを利用してこの5'領域に翻訳開始ATGが存在するか否か、およびその存在位置の検討を行った。その結果、独自に開発したソフトが翻訳開始コドンの存在の有無および存在位置を的確に判定することができ、このソフトにより翻訳開始コドンが存在すると判定されたクローンをcDNAライブラリーから選択することにより効率的に完全長のcDNAクローンを選択することが可能であることを見いだした。さらに、選択したクローンを混合することにより、完全長率の極めて高いcDNAライブラリーを作製できることを見いだした。

PCTに基づいて公開される国際出願のパンフレット第一頁に掲載されたPCT加盟国を同定するために使用されるコード(参考情報)

AE	アラブ首長国連邦	ES	スペイン	LI	リヒテンシュタイン	SG	シンガポール
AL	アルバニア	FI	フィンランド	LK	スリ・ランカ	SI	スロヴェニア
AM	アルメニア	FR	フランス	LR	リベリア	SK	スロヴァキア
AT	オーストリア	GA	ガボン	LS	レソト	SL	シエラ・レオネ
AU	オーストラリア	GB	英国	LT	リトアニア	SN	セネガル
AZ	アゼルバイジャン	GD	グレナダ	LU	ルクセンブルグ	SZ	スワジランド
BA	ボスニア・ヘルツェゴビナ	GE	グルジア	LV	ラトヴィア	TD	チャード
BB	バルバドス	GH	ガーナ	MC	モナコ	TG	トーゴ
BE	ベルギー	GM	ガンビア	MD	モルドヴァ	TJ	タジキスタン
BF	ブルキナ・ファソ	GN	ギニア	MG	マダガスカル	TM	トルクメニスタン
BG	ブルガリア	GW	ギニア・ビサオ	MK	マケドニア旧ユーゴスラヴィア	TR	トルコ
BJ	ベナン	GR	ギリシャ		共和国	TT	トリニダード・トバゴ
BR	ブラジル	HR	クロアチア	ML	マリ	UA	ウクライナ
BY	ベラルーシ	HU	ハンガリー	MN	モンゴル	UG	ウガンダ
CA	カナダ	ID	インドネシア	MR	モーリタニア	US	米国
CF	中央アフリカ	IE	アイルランド	MW	マラウイ	UZ	ウズベキスタン
CG	コンゴ	IL	イスラエル	MX	メキシコ	VN	ヴェトナム
CH	スイス	IN	インド	NE	ニジェール	YU	ユーゴスラビア
CI	コートジボアール	IS	アイスランド	NL	オランダ	ZA	南アフリカ共和国
CM	カメルーン	IT	イタリア	NO	ノルウェー	ZW	ジンバブエ
CN	中国	JP	日本	NZ	ニュージーランド		
CY	キプロス	KE	ケニア	PL	ポーランド		
CZ	チェコ	KG	キルギスタン	PT	ポルトガル		
DE	ドイツ	KP	北朝鮮	RO	ルーマニア		
DK	デンマーク	KR	韓国	RU	ロシア		
EE	エストニア	KZ	カザフスタン	SD	スーダン		
		LC	セントルシア	SE	スウェーデン		

## 明細書

### 完全長 cDNA クローンの選択方法

#### 技術分野

本発明は、遺伝子工学の分野に属し、完全長cDNAクローンを選択する方法に関する。

#### 背景技術

現在、種々の動植物および微生物においてゲノムプロジェクトが進展し、多くの遺伝子が単離されその機能の解析が試みられているが、単離した遺伝子の機能を効率よく解析するためには完全なタンパク質を発現可能なcDNAクローン、即ち、完全長のcDNAクローンを効率よく得ることが重要なステップとなる。

完全長率の高いcDNAライブラリーの作製法としては、現在のところ次のような方法が知られている。即ち、mRNAのCAPにRNAリンカーを酵素的に結合するオリゴキャップ法(菅野・丸山, 蛋白質 核酸 酵素, 38, 476 -481 (1993).、鈴木・菅野, 蛋白質 核酸 酵素, 41, 603-607 (1996).、M. Maruyama and S. Sugano, Gene, 138, 171-174 (1994).)、オリゴキャップ法とOkayama-Berg法を組み合わせで開発した改良オリゴキャップ法(S. Kato et al., Gene, 150, 243-250 (1994).、加藤・関根 特開平6-153953号公報 1994年6月3日)、CAPにDNAリンカーを化学結合するリンカー化学結合法(N. Merenkova and D. M. Edwards, WO 96/34981 Nov. 7, 1996.)、CAPのビオチン修飾によるキャップ化学修飾法(P. Carninci et al., Genomics, 37,327-336 (1996).、P. Carninci et al., DNA Research, 4, 61-66 (1997).)である。これらはすべて真核生物のmRNAのCAPを修飾する方法であり、完全長率の高いcDNAライブラリーを作製する方法である。また、Capをトラップすることにより完全長cDNAの比率の高いライブラリーを作製する方法としては、5'

Capサイトの標識法として、酵母またはHela細胞のCap結合タンパク質を用いる方法 (I. Edery et al., MCB, 15, 3363-3371 (1995)) などが知られている。これら以外の方法としては、第一鎖cDNAの5' 端をC-テイリングしてCap Switch oligonucleotideをアニールさせるCap Switch oligonucleotide法であるCap Finder (Clontech社製) が知られている。

しかしながら、従来の方法に比して完全長率が高いとはいえ、これら方法を用いてcDNAライブラリーを作製した場合には、不完全長のクローンが必ずある頻度で混入してしまう。そこで、遺伝子の高効率機能解析を行い、新規な有用遺伝子を高効率でクローン化するために、cDNAライブラリーに含まれる各クローンが完全長であるか否かを容易に識別しうる方法の開発が望まれていた。

#### 発明の開示

本発明は、効率的に完全長のcDNAクローンを選択する方法、および完全長率の高いcDNAライブラリーの作製方法を提供することを課題とする。

本発明者等は、上記課題を解決すべく鋭意研究を行った結果、5' 末端の一定領域が欠失したcDNAには翻訳開始コドンが含まれない可能性が高く、一方、完全長cDNAには翻訳開始コドンが存在することから、翻訳開始コドンの存在の有無を指標として、cDNAライブラリーから効率的に完全長cDNAを選択することができると考えた。特に、完全長率の高いcDNAライブラリー作成法で作製したcDNAライブラリーを用いれば、高効率で完全長cDNAを選択しうると考えた。即ち、完全長率の高いcDNAライブラリー作成法でcDNAライブラリーを作成後、5' 末端より数百塩基のDNA 塩基配列を決定し、この領域内に翻訳開始コドンが存在するか否かを解析して翻訳開始コドンが存在するクローンを選択することにより、効率的に完全長のcDNAクローンを単離することができると考えた。

しかしながら、翻訳開始コドンを予測する方法に関し、cDNAの翻訳開始点を予測するプログラムに関しては、ほとんど開発がなされていないのが現状である

(報告例としては、「A. G. Pedersen, Proceedings of fifth international conference on intelligent systems for molecular biology, p226-233 (1997), held in Halkidiki, Greece, June 21-26, 1997.」が挙げられる)。エキソンを予測するプログラムについては開発されてきてはいるが(「Gene Finder」V. V. Solovyev et al., Nucleic Acids Res., 22, 5156-5163 (1994).、「Grail」Y. Xu et al., Genet-Eng-N-Y., 16, 241-253 (1994))、このプログラムを用いることのみによって完全に翻訳開始点を決定することはできない。

そこで、本発明者等は、独自にcDNAの翻訳開始コドンの予測ソフトを開発した。そして、完全長率の高いcDNAライブラリー作成法で作製したcDNAライブラリーに含まれるクローンの5'領域の塩基配列を決定し、この開発したソフトを利用してこの5'領域に翻訳開始コドンが存在するか否かの検討を行った。

具体的には、まず、オリゴキャップ法で全長cDNAライブラリーを作製し、cDNAライブラリーに含まれるいくつかのクローンについてその5'末端側の塩基配列を決定し、決定した塩基配列に基づき、クローンをデータベース上で既知であるクローンと新規なクローンとに分別した。次いで、翻訳開始コドンの予測ソフトを用いて5'末端側の決定された塩基配列につき翻訳開始コドンの有無およびその存在位置を判定し、既知のクローンについては予測ソフトで認定された翻訳開始コドンと、データベース上の翻訳開始コドンの存在位置との整合性を検討した。その結果、本発明者等は、検討を行った既知のクローンにつき、予測ソフトにより判定された翻訳開始コドンの有無および存在位置が実際のデータベース上で事実と一致することを見いだした。

ここに本発明者等は、独自に開発したソフトが翻訳開始コドンの存在の有無および存在位置を的確に判定することができ、このソフトにより翻訳開始コドンが存在すると判定されたクローンをcDNAライブラリーから選択すれば、効率的に完全長のcDNAクローンを選択することが可能であることを見いだした。さらに、本発明者等は、選択したクローンを混合することにより、完全長率の極めて高いcD

NAライブラリーを作製できることを見いだした。

即ち、本発明は、cDNAライブラリーから完全長cDNAクローンを選択する方法、およびこれにより選択したcDNAクローンを混合することを特徴とする完全長cDNAライブラリーの作製法に関し、より具体的には、

(1) 完全長cDNAクローンを単離する方法であって、

(a) cDNAライブラリーに含まれるcDNAクローンの5'末端から塩基配列を決定する工程、

(b) (a)で決定された塩基配列中に開始コドンが含まれているか否かを、開始コドンを予測するプログラムによって判定する工程、

(c) (b)で開始コドンが含まれていると判定されたクローンを選択する工程、を含む方法、

(2) cDNAライブラリーが完全長率の高いcDNAライブラリー作製法により作製されたものである、(1)に記載の方法、

(3) cDNAライブラリーがmRNAのCAPを修飾する工程を含む方法により作製されたものである、(1)に記載の方法、

(4) 完全長cDNAライブラリーを作製する方法であって、

(a) cDNAライブラリーに含まれるcDNAクローンの5'末端から塩基配列を決定する工程、

(b) (a)で決定された塩基配列中に開始コドンが含まれているか否かを、開始コドンを予測するプログラムによって判定する工程、

(c) (b)で開始コドンが含まれていると判定されたクローンを選択する工程、

(d) (c)で選択されたクローンを混合する工程、を含む方法、

(5) cDNAライブラリーが完全長率の高いcDNAライブラリー作製法により作製されたものである、(4)に記載の方法、

(6) cDNAライブラリーがmRNAのCAPを修飾する工程を含む方法により作製されたものである、(4)に記載の方法、



(7) (4)に記載の方法により作製しうるcDNAライブラリー、  
に関する。

本発明は、cDNAライブラリー、特に完全長率の高いcDNAライブラリーの5'領域の塩基配列を、翻訳開始コドンを高精度で予測するソフトにより解析することにより、効率的に全長cDNAクローンを単離することができ、さらに単離したcDNAクローンを混合することにより完全長率をさらに高めたcDNAライブラリーを作製することができるとの本発明者等による知見に基づく。従って、本発明の完全長cDNAクローンを選択する方法は、(a) cDNAライブラリーに含まれるcDNAクローンの5'末端から塩基配列を決定する工程、(b) 決定された塩基配列中に開始コドンが含まれているか否かを、開始コドン进行予測するプログラムによって判定する工程、および(c) 開始コドンが含まれていると判定されたクローンを選択する工程を含む。また、本発明の完全長cDNAライブラリーの作製法は、さらに(d) 選択されたクローンを混合する工程を含む。

本発明の方法において、5'末端領域の塩基配列の決定を行う「cDNAクローン」としては、特に制限はないが、クローンが完全長率の高くないライブラリーに由来する場合には、完全長率の高いライブラリーに由来する場合と比較して、結果として完全長cDNAの単離の効率が低下すると考えられる。このためcDNAクローンは、上記の完全長率の高いcDNAライブラリーの作成法、例えば、mRNAのCAPにRNAリンカーを酵素的に結合するオリゴキャップ法(菅野・丸山, 蛋白質 核酸 酵素, 38, 476-481 (1993).、鈴木・菅野, 蛋白質 核酸 酵素, 41, 603-607 (1996).、M. Maruyama and S. Sugano, Gene, 138, 171-174 (1994).)、オリゴキャップ法とOkayama-Berg法を組み合わせで開発した改良オリゴキャップ法(S. Kato et al., Gene, 150, 243-250 (1994).、加藤・関根 特開平6-153953 1994年6月3日)、CAPにDNAリンカーを化学結合するリンカー化学結合法(N. Merenkova and D. M. Edwards, WO 96/34981 Nov. 7, 1996.)、CAPのビオチン修飾によるキャップ化学修飾法(P. Carninci et al., Genomics, 37, 327-336 (1996).、P. Carninci et

al., DNA Research, 4, 61-66 (1997).)、酵母またはHela細胞のCap結合タンパク質を用いる方法 (I.Edery et al., MCB, 15, 3363-3371 (1995))、Cap Switch oligonucleotide法であるCap Finderにより作製されたライブラリーに由来することが好ましい。

cDNAライブラリーからのcDNAクローンの単離は、文献 (J.Sambrook, E.F.Fritsch & T.Maniatis, Molecular Cloning Second edition, Cold Spring Harbor Laboratory Press 1989) などに記載の常法により行うことができる。

クローンの5'末端からの塩基配列の決定は、例えば、アプライドバイオシステムズ社のDNAシーケンス試薬とDNAシーケンサー等を用いて常法により行うことができる。塩基配列は全配列を決定する必要はなく、5'末端から1000塩基程度を決定すれば十分であるが、500塩基程度、さらには300塩基程度の塩基配列の決定であっても高い精度が期待できる。

クローンの5'末端からの塩基配列の解析に用いられる「開始コドンを予測するプログラム」としては、本発明者らにより開発された後述する実施例1に記載のプログラムが好ましい。決定された塩基配列中に開始コドンが含まれているか否かの判定は、プログラムによる解析の結果、抽出されるスコアにより行われる。スコアが高い値を示し決定した塩基配列中に開始コドンが存在すると判定されたcDNAクローンは、通常は、完全長cDNAであり、逆に、スコアが低い値を示し塩基配列中に開始コドンが存在しないと判定されたcDNAクローンは、不完全長cDNAである。従って、cDNAライブラリーから塩基配列中に開始コドンが存在すると判定されたcDNAクローンを選択することにより、効率的に完全長cDNAを単離することが可能である。実際に実施例1に記載のプログラムを用いた解析の一例では、完全長率が51%のcDNAライブラリーをクローンの選択に用いた場合、スコア（最高値0.94）を0.5以上でクローンを選択するとクローンの完全長率は71%、スコアを0.70以上でクローンを選択するとクローンの完全長率は77%、スコアを0.80以上でクローンを選択するとクローンの完全長率は81%、スコアを0.90以上でクローンを選

択するとクローンの完全長率は85%であった。従って、実施例1に記載のプログラムを用いてスコアの高いクローンを選択することにより、高い確率で完全長cDNAクローンが選択される。

また、本発明の完全長cDNAクローンの選択方法により選択されたクローンを混合することにより、クローンの選択に用いたcDNAライブラリーの完全長率を飛躍的に高めたcDNAライブラリーを再構築することも可能である。このようなタンパク質に発現可能なcDNAからなるcDNAライブラリーをライブラリーのまま発現させれば、混合物としてタンパク質を発現させた高効率遺伝子機能解析系を構築することができ、これにより有用な遺伝子を高効率でクローニングすることが可能となる。

#### 発明を実施するための最良の形態

以下、本発明を実施例によりさらに詳細に説明するが本発明はこれら実施例に制限されるものではない。

##### 〔実施例1〕 cDNAの翻訳開始コドン进行予測するプログラムの作製

本発明のcDNAの翻訳開始コドン进行予測するプログラムは、cDNA配列断片が与えられた時に、その中に含まれる全てのATGの中から真の翻訳開始コドンらしいものを予測する。予測は、基本的には、A) 着目したATGの両側の一定範囲領域（数10～数100塩基）のそれぞれにおける翻訳領域らしさの情報と、B) 着目したATGの付近の翻訳開始コドンらしさの情報を用いて行う。あらかじめ、翻訳領域と非翻訳領域が判明している配列を多数用いて、翻訳領域と翻訳開始コドン付近の配列の特徴を抽出しておく。プログラムは、これらの特徴情報を用いて翻訳開始コドン进行予測する。

ゲノムエクソン予測プログラムであるGene Finder (Solovyev V.V., Salamov A.A., Lawrence C.B. Predicting internal exons by oligonucleotide composition and discriminant analysis of spliceable open reading frames. Nucleic

Acids Res.(1994) 22: 5156-63.) 中で用いられている線形判別分析を、予測の最適化の手法として用いた。線形判別分析では、データの持つ複数の特徴情報をそれぞれ数量化し、それらに重みをつけて加算した量をスコアとして用いた。ここでは、スコアを翻訳開始コドンらしさの確率（翻訳開始コドンがすでに判明しているデーターを用いた場合の正解率を確率とする）に変換する。即ち、与えられたcDNA配列断片に含まれる各ATGに対して翻訳開始コドンらしさの確率が出力される。翻訳開始コドンの認定は、翻訳開始コドンらしさの確率が一定のしきい値を超えるかどうかで行うが、しきい値はその後の解析の方針に応じて、即ち、どの程度のノイズを許容して解析を行うかに応じて、その値を設定する。例えば、40%のノイズを許容可能であれば、しきい値0.6を採用すればよい。重みのパラメーターは、翻訳開始コドンがすでに判明しているデーターをトレーニングデーターとして用いて、予測精度が最大になるように決定する。使用した特徴情報は、上述のAとBの情報を具体化したそれぞれ3つの情報を用いた。

具体的には、A) 着目したATGの両側の一定範囲領域（数10～数100塩基）のそれぞれにおける翻訳領域らしさの情報としては、①ATGからストップコドンまで（最大で、ATGの300塩基後まで）に含まれる6塩基文字の頻度情報、②ATGの前後50塩基に含まれる6塩基文字の頻度情報の差、③シグナルペプチドらしさの指標（ATGの後30アミノ酸（90塩基）の中で、最も疎水的な8アミノ酸文字の疎水性指標）を用いた。また、B) 着目したATGの付近の翻訳開始コドンらしさの情報としては、①ATGの14塩基前から5塩基後までの領域内の3塩基文字を単位とした重み行列情報、②ATGの前に同じフレームで他のATGがあるかないか（あれば1、なければ0）、③ATGの36塩基前から7塩基前までの領域に含まれるシトシン塩基の頻度を用いた。

【実施例2】 オリゴキャップ法によるcDNAの調製、および翻訳開始コドン予測プログラム（ATGpr）による解析

オリゴキャップ法でcDNAライブラリーを作成し、個々のクローンについて常法

によりプラスミドDNAを抽出した。具体的には、ヒト胎盤およびヒト培養細胞 (teratocarcinoma NT-2、neuroblastoma SK-N-MC) より文献 (J.Sambrook, E.F.Fritsch & T.Maniatis, Molecular Cloning Second edition, Cold Spring Harbor Laboratory Press 1989) 記載の方法により mRNA を抽出した。次いで、オリゴキャップリンカー (配列番号: 1) と、オリゴdTアダプタープライマー (配列番号: 2) (表 1 と 2 の場合)、またはランダムアダプタープライマー (配列番号: 3) (表 3 と表 4 の場合) を用いて文献 (鈴木・菅野, 蛋白質 核酸 酵素, 41, 603-607 (1996). p606、Y.Suzuki et al., Gene, 200, 149-156 (1997)) の記載に従い、BAP処理、TAP処理、RNAライゲーション、第一鎖cDNAの合成とRNAの除去を行った。次いで、PCRにより2本鎖cDNAに変換し、SfiI切断した。次いで、DraIIIで切断したベクターpME18SCG、pMFL等 (菅野・丸山, 蛋白質 核酸 酵素, 38, 472-481 (1993). p480) にcDNAの方向性を決めてクローニングした。これにより得たDNAについてDNAシーケンシング試薬(Dye Terminator Cycle Sequencing FS Ready Reaction Kit, PE Applied Biosystems)を用いてマニュアルにしたがってシーケンシング反応後、DNAシーケンサー(ABI PRISM 377, PE Applied Biosystems)でDNA塩基配列を解析した。各クローンの5' 末より、それぞれ1回のみDNA配列を解析した。

各クローンのDNA配列について、開発したcDNAの翻訳開始コドン予測プログラム(ATGpr)で翻訳開始コドンが存在するか否かにつき解析した。この解析プログラムでは数値が高いほど翻訳開始コドンである確率が高いことを示す。ただし、最高値は0.94である。

(1) オリゴキャップ法でcDNAを作成したcDNAのうち、すでにオープンリーディングフレームがデータベースで既知のクローンについての翻訳開始コドンの解析

解析を行ったクローンのうち、決定した配列中に翻訳開始コドンが存在することがデータベースで既知のクローン (F-NT2RP1000020、F-NT2RP1000025、F-NT2RP1000039、F-NT2RP1000046) についての解析結果を以下に示す (表 1)。F-NT2R

P1000020(880 bp)は、88-690位が「Human neuron-specific gamma-2 enolase」(GenBank accession No.M22349)に対し相同性が96%、F-NT2RP1000025(645 bp)は、29-641位が「Human alpha-tubulin mRNA」(GenBank accession No.K00558)に対し相同性が97%、F-NT2RP1000039(820 bp)は、12-820位が「Human mRNA for elongation factor 1 alpha subunit(EF-1 alpha)」(GenBank accession No. X03558)に対し相同性が96%、F-NT2RP1000046(788 bp)は3-788位が「Human M2-type pyruvate kinase mRNA」(GenBank accession No.M23725)に対し相同性が97%である。なお、これらクローンの5'末端側の塩基配列をそれぞれ配列番号：4、5、6、7に示す。

表 1

F-NT2RP1000020			F-NT2RP1000025		F-NT2RP1000039		F-NT2RP1000046	
ATG No.	ATGの位置	ATGprスコア	ATGの位置	ATGprスコア	ATGの位置	ATGprスコア	ATGの位置	ATGprスコア
1	1	0.05	96	<0.94>	65	<0.90>	111	<0.94>
2	162	<0.84>	148	0.13	154	0.05	174	0.82
3	292	0.05	193	0.05	209	0.11	198	0.19
4	313	0.05	201	0.09	231	0.05	300	0.16
5	441	0.05	232	0.05	321	0.05	315	0.11

(注1) < > : 翻訳開始コドン

(注2) ATGの位置：5'-末よりのDNA配列でのATGの存在する塩基No.を示す。

ATG No. : 5'-末よりのDNA配列でのATGのNo.を示す。

表1が示すように、オリゴキャップ法でcDNAを作成したcDNAのうち、すでにオ

オープンリーディングフレームがデータベースで既知で翻訳開始コドンをもつ完全長のクローンについて開発したcDNAの翻訳開始コドン予測プログラム (ATGpr) で解析した結果、翻訳開始コドンを間違いなく同定できた (データベース上の翻訳開始コドンと一致した)。

(2) オリゴキャップ法でcDNAを作成したcDNAのうち、翻訳開始コドンが存在しないことがデータベースで既知のクローンについての翻訳開始コドンの解析解析を行ったクローンのうち、決定した配列中に翻訳開始コドンが存在しないことがデータベースで既知のクローン (F-NT2RP1000013、F-NT2RP1000054、F-NT2RP1000122) についての解析結果を以下に示す (表2)。F-NT2RP1000013(608 bp)は1-606位が「Human nuclear matrix protein 55 (nmt55) mRNA」(GenBank accession No. U89867)に対し相同性が97%、F-NT2RP1000054(869 bp)は1-869位が「Human signal recognition particle (SRP54) mRNA」(GenBank accession No. U51920)に対し相同性が96%、F-NT2RP1000122(813 bp)は1-813位が「H.sapiens mRNA for 2-5A binding protein」(GenBank accession No.X76388)に対し相同性が98%であった。なお、これらクローンの5'末端側の塩基配列をそれぞれ配列番号: 8、9、10に示す。

表 2

F-NT2RP1000013			F-NT2RP1000054		F-NT2RP1000122	
ATG	ATGの	ATGpr	ATGの	ATGpr	ATGの	ATGpr
No.	位置	スコア	位置	スコア	位置	スコア
1	21	0.05	31	0.12	23	0.07
2	27	0.05	60	0.20	100	0.05
3	32	0.32	87	0.05	166	0.05
4	56	0.11	97	0.05	235	0.06
5	119	0.10	146	0.05	316	0.05
6	125	0.08	172	0.05	346	0.05
7	141	0.05	180	0.11	406	0.05
8	155	0.06	218	0.07	431	0.05
9	161	0.06	272	0.05	469	0.06
10	176	0.08	319	0.07	546	0.12
11	203	0.07	346	0.05	553	0.05
12	290	0.20	363	0.07	574	0.05
13	311	0.16	409	0.05		
14	314	0.12	480	0.07		

表 2 が示すようにオリゴキャップ法で cDNA を作成した cDNA のうち、すでにオープンリーディングフレームがデータベースで既知で翻訳開始コドンをもたない不完全長のクローンについて開発した cDNA の翻訳開始コドン予測プログラム (ATGpr) で解析した結果、翻訳開始コドンを間違えて同定することはなかった。

(3) オリゴキャップ法で cDNA を作成した cDNA のうち、新規配列のクローンに



ついでに翻訳開始コドンの解析

解析を行ったクローンのうち、翻訳開始コドンが存在すると予測された新規クローン（F-ZRV6C1000408、F-ZRV6C1000454、F-ZRV6C1000466、F-ZRV6C1000615、F-ZRV6C1000670）についての解析結果を以下に示す（表3）。なお、これらクローンの5'末端側の塩基配列をそれぞれ配列番号：11、12、13、14、15に示す。

表 3

F-ZRV6C1000408			F-ZRV6C1000454		F-ZRV6C1000466	
ATG	ATGの	ATGpr	ATGの	ATGpr	ATGの	ATGpr
No.	位置	スコア	位置	スコア	位置	スコア
1	85	<0.94>	5	0.05	162	<0.86>
2	208	0.22	107	<0.87>	182	0.05
3	386	0.05	153	0.05	207	0.08
4	518	0.11	201	0.08	244	0.05
5	545	0.05	211	0.05	262	0.05
6			236	0.07	303	0.11

F-ZRV6C1000615			F-ZRV6C1000670	
ATG	ATGの	ATGpr	ATGの	ATGpr
No.	位置	スコア	位置	スコア
1	85	<0.94>	120	<0.94>
2	208	0.26	187	0.54
3	386	0.05	312	0.06
4	518	0.09	388	0.05
5	545	0.05	445	0.05

(注) < > : 翻訳開始コドンと予測される

表 3 が示すように F-ZRV6C1000408、F-ZRV6C1000454、F-ZRV6C1000466、F-ZRV6C1000615、F-ZRV6C1000670 については、それぞれ 85 位、107 位、162 位、85 位、12

0位の塩基「A」から始まる「ATG」が開始コドンであると判定された。このためこれらクローンは完全長cDNAクローンであると判断した。

また、解析を行ったクローンのうち、翻訳開始コドンが存在しないと予測された新規クローン（F-ZRV6C1001410、F-ZRV6C1001197、F-ZRV6C1001472）についての解析結果を以下に示す（表4）。なお、これらクローンの5'末端側の塩基配列をそれぞれ配列番号：16、17、18に示す。

表4

F-ZRV6C1001410			F-ZRV6C1001197		F-ZRV6C1001472	
ATG	ATGの	ATGpr	ATGの	ATGpr	ATGの	ATGpr
No.	位置	スコア	位置	スコア	位置	スコア
1	23	0.05	5	0.24	77	0.25
2	31	0.07	141	0.25	126	0.05
3	71	0.06	202	0.05	149	0.05
4	178	0.05	219	0.05	194	0.05
5	214	0.05	228	0.05	213	0.22
6					249	0.05
7					338	0.09
8					344	0.05
9					351	0.05
10					365	0.05

表4が示すようにF-ZRV6C1001410、F-ZRV6C1001197、F-ZRV6C1001472については、開始コドンが存在しないと判定された。このためこれらクローンは不完全長

クローンであると判断した。

#### 産業上の利用の可能性

本発明により、効率的に完全長のcDNAクローンを選択する方法が提供された。本発明の方法により選択されたクローンは、完全なタンパク質を発現させることが可能である。このため本発明により、単離した遺伝子の機能解析を効率的に行うことが可能となった。

## 請求の範囲

1. 完全長cDNAクローンを選択する方法であって、
  - (a) cDNAライブラリーに含まれるcDNAクローンの5'末端から塩基配列を決定する工程、
  - (b) (a)で決定された塩基配列中に開始コドンが含まれているか否かを、開始コドンを予測するプログラムによって判定する工程、
  - (c) (b)で開始コドンが含まれていると判定されたクローンを選択する工程、を含む方法。
2. cDNAライブラリーが完全長率の高いcDNAライブラリー作製法により作製されたものである、請求項1に記載の方法。
3. cDNAライブラリーがmRNAのCAPを修飾する工程を含む方法により作製されたものである、請求項1に記載の方法。
4. 完全長cDNAライブラリーを作製する方法であって、
  - (a) cDNAライブラリーに含まれるcDNAクローンの5'末端から塩基配列を決定する工程、
  - (b) (a)で決定された塩基配列中に開始コドンが含まれているか否かを、開始コドンを予測するプログラムによって判定する工程、
  - (c) (b)で開始コドンが含まれていると判定されたクローンを選択する工程、
  - (d) (c)で選択されたクローンを混合する工程、を含む方法。
5. cDNAライブラリーが完全長率の高いcDNAライブラリー作製法により作製されたものである、請求項4に記載の方法。
6. cDNAライブラリーがmRNAのCAPを修飾する工程を含む方法により作製されたものである、請求項4に記載の方法。
7. 請求項4に記載の方法により作製しうるcDNAライブラリー。

## 配列表

## SEQUENCE LISTING

<110> Helix Research Institute, Inc.

株式会社ヘリックス研究所

<120> The method for selecting full length cDNA clones

完全長 cDNA クローンの選択方法

<130> H1-806PCT

<150> JP 09-289982

<151> 1997-10-22

<160> 18

<170> PatentIn version 2.0

<210> 1

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> Oligo-capping linker sequence

2/14

&lt;400&gt; 1

AGCAUCGAGU CGGCCUUGUU GGCCUACUGG

30

&lt;210&gt; 2

&lt;211&gt; 42

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Oligo(dT) adapter primer sequence

&lt;400&gt; 2

GCGGCTGAAG ACGGCCTATG TGGCCTTTTT TTTTTTTTTT TT

42

&lt;210&gt; 3

&lt;211&gt; 32

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Random adapter primer sequence

&lt;400&gt; 3

GCGGCTGAAG ACGGCCTATG TGGCCNNNNN NC

32

3/14

&lt;210&gt; 4

&lt;211&gt; 880

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 4

ATGCGCCCGC GCGGCCCTAT AGGCGCCTCC TCCGCCCGCC GCCCGGGAGC CGCAGCCGCC 60  
GCCGCCACTG CCACTCCCGC TCTCTCAGCG CCGCCGTCGC CACCGCCACC GCCACTGCCA 120  
CTACCACCGT CTGAGTCTGC AGTCCCGAGA TCCCAGCCAT CATGTCCATA GAGAAGATCT 180  
GGGCCCCGGA GATCCTGGAC TCCCGCGGGA ACCCCACAGT GGAGGTGGAT CTCTATACTG 240  
CCAAAGGTCC TTTCCGGGCT GCAGTGCCCA GTGGAGCCTC TACGGGCATC TATGAGGCCC 300  
TGGAGCTGAG GGATGGAGAC AAACAGCGTT ACTTAGGCAA AGGTGTCCTG AAGGCAGTGG 360  
ACCACATCAA CTCCACCATC GCGCCAGCCC TCATCAGCTC AGGTCTCTCT GTGGTGGAGC 420  
AAGAGAAACT GGACAACCTG ATGCTGGAGT TGGATGGGAC TGAGAACAAA TCCAAGTTTG 480  
GGGCCAATCC ATCCTGGGTG TGTCTCTGGC CGTGTGTAAG GCANGGGCAA CTGAACNGGA 540  
ACTGCCCTG TATCGCCACA TTGCTCAGCT TGGNCGGGAA CTCANACCTC ATCCTGCCTG 600  
TTGCCGGCCT TCAACGTGAT CAATGGTTGG CTTCTCATGC CTGGCAACAA ANCTGGCCAT 660  
TGCNGGAATT TTCATGATCC TCCCNTTGG GAAACTGAAA AACTTTCCGG AATGCCCNTC 720  
CAACTAAGTT GCAAAAGGTC TACCNATACC CCCCAGGGG AATTCCTCCA AGGGAACAAA 780  
-----  
TNCCCGGGAA AGGAATGCCC CCAATTNTT NGGGGAATA AAAGGTGGGC TTTGCCCCC 840  
CATTTTCCTG GAAAAACNA TNA AACCT TGGGAACTT 880

&lt;210&gt; 5

&lt;211&gt; 645



4/14

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 5

TGTGCGTTAC TTACCTCNAC TCTTAGCTTG TCGGGGACGG TAACCGGGAC CCGGTGTCTG 60  
CTCCTGTCGC CTTGCGCTCC TAATCCCTAG CCACTATGCG TGAGTGCATC TCCATCCACG 120  
TTGGCCAGGC TGGTGTCCAN ATTGGCAATG CCTGCTGGGA GCTCTACTGC CTGGAACACG 180  
GCATCCAGCC CGATGGCCAG ATGCCAAGTG ACAAGACCAT TGGGGGAGGA GATGACTCCT 240  
TCAACACCTT CTTCAGTGAG ACGGGCGCTG GCAANCACGT GCCCCGGGCT GTGTTTGTAG 300  
ACTTGGAACC CACAGTCATT GATGAAGTTC GCACTGGCAC CTACCGCCAG CTCTTCCACC 360  
CTGAGCAGCT CATCNCAGGC AAGGAAGATG CTGCCAATAA CTATGCCCGA GGGCACTACA 420  
CCATTGGCAA GGAGATCATT GACCTTGTGT TGGACCGAAT TCGCAAGCTG GCTGACCANT 480  
GCACCGGTCT TCANGGCTTC TTGGTTTCC ACAGCTTTGG TGGGGGAACT GGTTCCTGGGT 540  
TCACCTCCCT GCTCATGGAA CGTCTCTCAG TTGATTATGG CAAGAAATCC AAGCTGGAGT 600  
TCTCCATTTA CCCAGCACCC CNGGTTTCCN CNGCTGTANT TNGAA 645

&lt;210&gt; 6

&lt;211&gt; 820

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 6

CTTTTTTCGC AACGGGTTTG CCGCCAGAAC ACAGGTGTCTG TGAAACTAC CCCTAAAAGC 60  
CAAAATGGGA AAGGAAAAGA CTCATATCAA CATTGTCGTC ATTGGACACG TAGATTCTGGG 120  
CAAGTCCACC ACTACTGGCC ATCTGATCTA TAAATGCGGT GGCATCGACA AAAGAACCAT 180

5/14

TGAAAAATTT GAGAAGGAGG CTGCTGAGAT GGGAAAGGGC TCCTTCAAGT ATGCCTGGGT 240  
CTTGGATAAA CTGAAAGCTG AGCGTGAACG TGGTATCACC ATTGATATCT CCTTGTGGAA 300  
ATTTGAGACC AGCAAGTACT ATGTGACTAT CATTGATGCC CCAGGACACA GAGACTTTAT 360  
CAAAAACATG ATTACAGGGA CATCTCAGGC TGA CTGTGCT GTCCTGATTG TTGCTGCTGG 420  
TGTTGGTGAA TTTGAAGCTG GTATCTCCAA GAATGGGCAG ACCCGAGAGC ATGCCCTTCT 480  
GGCTTACACA CTGGGTGTGA AACAACTAAT TGTCGGTGTT AACAAAATGG ATTCACTGAN 540  
CCACCCTACA GCCAGAAGAA ATATGANGAA ATTGTTAAGG AAGTCAGCAC TTACATTAAG 600  
AAAATTGGCT ACAACCCCGA CACAGTANCA TTTGTGCCAA TTTCTGGTTG GAATGGTGAC 660  
AACATGCTGG AACCAANTGC TAACATGCCT TGGTTCCAGG GATGGAAAAT CCCCCNTTAA 720  
GGATGGCNAT GCCATTGGAA CCCCCCTGCT TGAAGGCTCT GGANTGCATC CTANCACCAA 780  
CTCCTTCAAA TTGAAAAACC CCTTGCNCCC GCCTCCNCCA 840

&lt;210&gt; 7

&lt;211&gt; 788

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 7

---

GAGGCTGAGG CAGTGGCTCC TTGCACAGCA GCTGCACGCG CCGTGGCTCC GGATCTCTTC 60  
GTCTTTGCAG CGTAGCCCGA GTCGGTCAGC GCCGGAGGAC CTCAGCAGCC ATGTCGAAGC 120  
CCCATAGTGA AGCCGGGACT GCCTTCATTC AGACCCAGCA GCTGCACGCA GCCATGGCTG 180  
ACACATTCTT GGAGCACATG TGCCGCCTGG ACATTGATTC ACCACCCATC ACAGCCCGGA 240  
ACACTGGCAT CATCTGTACC ATTGGCCCAG CTTCCTGATC AGTGGAGACG TTGAAGGAGA 300  
TGATTAAGTC TGGAATGAAT GTGGCTCGTC TGAACCTCTC TCATGGAACT CATGAGTACC 360  
ATGCGGAGAC CATCAAGAAT GTGCGCACAG CCACGGAAAG CTTTGCTTCT GACCCCATCC 420

6/14

TCTACCGGCC CGTTGCTGTG GCTCTAGACA CTAAAGGACC TGAGATCCGA ACTGGGCTCA 480  
TCAAGGGCAG CGGCACTGCA GAGGTGGAGC TGAAGAATGG AGCCACTCTC AAAATCACGC 540  
TGGATAATGC CTACATGGAA AAGTGTGACG AGAACATCCT GTGGCTGGAC TACAAGAACA 600  
TCTGCAAGGT GGTGGAAGTG GGCAACAAGA TCTACGTGGA TGATGGGCTN ATTTCTCTCC 660  
AGGTGAACAC AAAGGTGCCG ACTTCCTGGG TGACNGANGT GGAAAATGGT GGCTCCTTGG 720  
GCNCAAGAAA GGTGTGAACT TCCTGGGGCT GCTGTGGANT TGCCTGCTGT GTCNGAAAAA 780  
GACATCCA 788

&lt;210&gt; 8

&lt;211&gt; 608

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 8

ACAGCCTGGC TCCTTTGAGT ATGAATATGC CATGCGCTGG AAGGCACTCA TTGAGATGGA 60  
GAAGCAGCAG CAGGACCAAG TGGACCGCAA CATCNAGGAG GCTCGTGAGA AGCTGGAGAT 120  
GGAGATGGAA GCTGCACGCC ATGAGCACCA GGTCATGCTA ATGAGACAGG ATTTGATGAG 180  
GCGCCAAGAA GAACTTCGGA GGATGGAAGA GCTGCACAAC CAAGANGTGC AAAAACGAAA 240  
GCAACTGGAG CTCAGGCAGG AGGAANAGCG CAGGCGCCGT GAAGAAANAGA TCGGCGGCA 300

---

GCAAGAAGAA ATGATGCGGC GACNGCAGGA AGGATTCAAG GGAACCTTCC CTGATGCGAG 360  
AGAGCAGGAG ATTCGGATGG GTCNGATGGC TATGGGAGGT GCTATGGGCA TAAACNACAG 420  
ATGTGCCATG CCCCCTGCTC CTGTGCCAGC TGGTACCCCA GCTCCTCCAG GACCTGCCAC 480  
TATTATGCCG GATGGAACCT TGGGATTGAC CCCACCNACA ACTGAACGCT TTGGTCNGGC 540  
TGCTACNATG GAANGAATTG GGGCAATTGG TGGAACCTCCT CCTGCATTCTN ACCGTGCAGC 600  
TCCTGGGA 608

7/14

&lt;210&gt; 9

&lt;211&gt; 869

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 9

```
ATATTAACT AGTGAAGCAA CTAAGAGAAA ATGTTAAGTC TGCTATTGAT CTTGAAGAGA 60
TGGCATCTGG TCTTAACAAA AGAAAAATGA TTCAGCATGC TGTATTTAAA GAACTTGTGA 120
AGCTTGTAGA CCCTGGAGTT AAGGCATGGA CACCCACTAA AGGAAAACAA AATGTGATTA 180
TGTTTGTTGG ATTGCAAGGG AGTGGTAAAA CAACAACATG TTCAAAGCTA GCATATTATT 240
ACCAGAGGAA AGGTTGGAAG ACCTGTTTAA TATGTGCAGA CACATTCAGA GCAGGGGCTT 300
TTGACCAACT AAAACAGAAT GCTACCAAAG CAAGAATTCC ATTTTATGGA AGCTATACAG 360
AAATGGATCC TGTCATCATT GCTTCTGAAG GAGTAGAGAA ATTTAAAAAT GAAAATTTTG 420
AAATTATTAT TGTTGATACA AGTGGCCGCC ACAAACAAGA AGACTCTTTG TTTGAAGAAA 480
TGCTTCAAGT TGCTAATGCT ATACAACCTG ATAACATTGT TTATGTGATG GATGCCTCCA 540
TTGGGCAGGC TTGTGAAGCC CAGGCTAAGG CTTTAAAGA TAAAGTAGAT GTACCTCAGT 600
AATAGTGACA AAACCTTGATG GCCATGCAAA ANGAAGTGGT GCACTCAGTG CAGTCGCTGC 660
CACAAAAAAT CCGATTATTT TCATTGGTAC AGGGGGAACA TATANATGAC TTTGAACCTT 720
-----
TCAAAAACAC AGCCTTTTAT TAACAAACTT CTTGGTATNG GCGACATTGA AAGGACTGAT 780
AAATAAAGTC CACNAATTGA AATTTGGATG ACNATGNAAA CCCTTATTGA AAAAATTGAA 840
ACATNGTCCA GTTTTACTTT GCGAAACNT 869
```

&lt;210&gt; 10

8/14

&lt;211&gt; 813

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 10

```
GTTGTGGTAT CTGTATTAAG AAATGCCCCT TTGGCGCCTT ATCAATTGTC AATCTACCAA 60
GCAACTTGGA AAAAGAAACC ACACATCGAT ATTGTGCCAA TGCCTTCAAA CTTCACAGGT 120
TGCCTATCCC TCGTCCAGGT GAAGTTTTGG GATTAGTTGG AACTAATGGT ATTGGAAAGT 180
CAACTGCTTT AAAAATTTTA GCAGGAAAAC AAAAGCCAAA CCTTGGAAAG TACGATGATC 240
CTCCTGACTG GCAGGAGATT TTGACTTATT TCCGTGGATC TGAATTACAA AATTACTTTA 300
CAAAGATTCT AGAAGATGAC CTAAGGCCA TCATCAAACC TCAATATGTA GACCAGATTC 360
CTAAGGCTGC AAAGGGGACA GTGGGATCTA TTTTGGACCG AAAAGATGAA ACAAAGACAC 420
AGGCAATTGT ATGTCAGCAG CTTGATTAA CCCACCTAAA AGAACGAAAT GTTGAAGATC 480
TTTCAGGAGG AGAGTTGCAG AGATTTGCTT GTGCTGTCGT TTGCATACAG AAAGCTGATA 540
TTTTCATGTT TGATGAGCCT TCTAGTTACC TAGATGTCAA GCAGCGTTTA AAGGCTGCTA 600
TTACTATACG ATCTCTAATA AATCCAGATA GATATATCAT TGTGGTGGAA CATGATCTAA 660
GTGTATTAGA CTATCTCTCC GACTTCATCT GCTGTTTATA TGGTGTACCA AGCGCCTATG 720
GAATTGTCAC TATGCCTTTT AGTGTTAGAA AAGGCATAAA CNTTTTTTGG ATGGGTATGT 780
TCCAACAGAA AACTTGANAA TCNNAAATGC NTC 813
```

&lt;210&gt; 11

&lt;211&gt; 655

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

9/14

&lt;400&gt; 11

AGACTCTCAC CGCAGCGGCC AGGAACGCCA GCCGTTACAG CGTTCGGTCC TCCTTGGCTG 60  
ACTCACC GCC CTGCGCGCCG CACCATGGAC GCCCCAGGC AGGTGGTCAA CTTTGGGCCT 120  
GGTCCCGCCA AGCTGCCGCA CTCAGTGTG TTAGAGATAC AAAAGGAATT ATTAGACTAC 180  
AAAGGAGTTG GCATTAGTGT TCTTGAAATG AGTCACAGGT CATCAGATTT TGCCAAGATT 240  
ATTAACAATA CAGAGAATCT TGTGCGGGAA TTGCTAGCTG TTCCAGACAA CTATAAGGTG 300  
ATTTTTCTGC AAGGAGGTGG GTGCGGCCAG TTCAGTGCTG TCCCCTTAAA CCTCATTGGC 360  
TTGAAAGCAG GAAGGTGTGC GGAATATGTG GTGACAGGAG CTTGGTCAGC TAAGGCCGCA 420  
GAAGAAGCCA AGAAGTTTGG GACTATAAAT ATCGTTCACC CTAAACTTGG GAGTTATACA 480  
AAAATTCCAG ATCCAAGCAC CTGGAACCTC AACCCANATG CCTCCTACGT GTTTTATTGC 540  
NCAAATGAAA CGGTGCATGG TGTTGANTTT GACTTTATAC CCNATGTCAA GGGAAACANTA 600  
CTGGTTTGTG ACATTTTCCT CCAACTTCCT GTCCAANCCA ATTGNATGTT TCCAA 655

&lt;210&gt; 12

&lt;211&gt; 599

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 12

---

AAAGATGCGC AGGCGCCGTG TGGCACTCGG CGGTCGAAAG GGGAGTTCAA GGAGACGGGG 60  
GCGACGCGGC TGAGGGCTTC TCGTCGGGGT CGGGGCTGCA GCCGTCATGC CGGGGATAGT 120  
GGAGCTGCCC ACTCTAGAGG AGCTGAAAGT AGATGAGGTG AAAATTAGTT CTGCTGTGCT 180  
TAAAGCTGCG GCCCATCACT ATGGAGCTCA ATGTGATAAG CCCAACAAGG AATTTATGCT 240  
CTGCCGCTGG GAANAGAAAG ATCCGAGGCG GTGCTTAGAG GAAGGCAAAC TGGTCAACAA 300  
GTGTGCTTTG GACTTCTTTA GGCAGATAAA ACGTCACTGT GCAGAGCCTT TTACAGAATA 360

10/14

TTGGACTTGC ATTGATTATA CTGGCCAGCA GTTATTTCGT CACTGTCGCA AACAGCAGGC 420  
AAAGTTTGAC NAGTGTGTGC TGGACAAACT GGGCTGGGTG CGGCCTGACC TGGGAAAACT 480  
GTCAAAGGTC ACCAAAGTGA AAACAGATCN ACCTTTACCG GANAATCCCT ATCACTCAAG 540  
AACAAGAACG GATCCCAGCC CTGANATCNA AGGAAATCTG CANCCTGCCA CACATGGCA 599

&lt;210&gt; 13

&lt;211&gt; 597

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 13

ATATCCGGAG TAGACGGAGC CGCAGTAGAC GGATCCGCGG CTGCACCAAA CACTGCCCCCT 60  
CGGAGCCTGG TAGTGGGCCA CAAGCCCCCA GTCCCAGAGG CGTGATTTTC TGGCATCCTT 120  
AAATCTTG TG TCAAGGATTG GTTATAATAT AACCAGAAAC CATGACGGCG GCTGAGAACG 180  
TATGCTACAC GTTAATTAAC GTGCCAATGG ATTCAGAACC ACCATCTGAA ATTAGCTTAA 240  
AAAATGATCT AGAAAAAGGA GATGTAAAGT CAAAGACTGA AGCTTTGAAG AAAGTAATCA 300  
TTATGATTCT GAATGGTGAA AACTTCCTG GACTTCTGAT GACCATCATT CGTTTTGTGC 360  
TACCTCTTCA GGATCACACT ATCAAGAAAT TACTTCTGGT ATTTTGGGAG ATTGTTCTTA 420  
AAACAACCTC AGATGGGAGA CTTTACATG AGATGATCCT TGTATGTGAT GCATACAGAA 480

---

AGGATCTTCA ACATCCTAAT GAATTTATTC NAAGGATCTA CTCTTCGTTT TCTTTGCAAA 540  
TTGAAANAAA CANAATTGCT AAAACCTTTA ATGCCANCTA TNCCTGCATT TTTGGGA 597

&lt;210&gt; 14

&lt;211&gt; 634

11/14

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 14

```
AGACTCTCAC CGCAGCGGCC AGGAACGCCA GCCGTTACAG CGTTCGGTCC TCCTTGGCTG   60
ACTCACCGCC CTCGCCGCCG CACCATGGAC GCCCCAGGC AGGTGGTCAA CTTTGGGCCT   120
GGTCCCGCCA AGCTGCCGCA CTCAGTGTTG TTAGAGATAC AAAAGGAATT ATTAGACTAC   180
AAAGGANTTG GCATTAGTGT TCTTGAAATG AGTCACAGGT CATCAGATTT TGCCAAGATT   240
ATTAACAATA CAGAGAATCT TGTGCGGGAA TTGCTAGCTG TTCCAGACAA CTATAAGGTG   300
ATTTTCTGC AAGGAGGTGG GTGCGGCCAG TTCAGTGCTG TCCCCTTAAA CCTCATTGGC   360
TTGAAAGCAG GAANGTGTGC GGACTATGTG GTGACAGGAG CTTGGTCAGC TAAGGCCGCA   420
NAANAAGCCA AGAANTTTGG GACTATAAAT ATCGTTCACC CTAAACTTGG GAGTTATACA   480
AAAATTCCAG ATCCAAGCAC CTGGAACCTC AACCAGATG CCTCCTACGT GTATTATTGC   540
GCNAATGAAA CNGTGCATGG TGTGGANTCT GACTTTATAC CCGATGTCNA GGGAACATAC   600
TGGTTTGTGA CATGTCCTCA AACTTCCCGT CCNA                               634
```

&lt;210&gt; 15

&lt;211&gt; 757

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 15

```
AGTCTGCGGT GGGCTANCGG ACGGTCCGGC TTCCGGCGGC CGTTTCTGTC TCTTGCTGGC   60
TGTCTCGCTG AATCGCGGCC GCCTTCTCAT CGCTCCTGGA AGGTCCCGAG CGCGACACCA   120
TGTCGGAACC CGGGGGCGGC GCGGCGAAG ACNGCTCGGC CGGATTGGAA GTGTCGGCCG   180
```



12/14

TGCANAATGT GCGGGACGTG TCGGTGCTGC ANAAGCACCT GCGCAAGCTG GTGCCGCTGC 240  
TGCTGGAGGA CGGCGGCGAA GCGCCGGCCG CGCTGGAGGC GCGCTGGAG GAGAAGAGCG 300  
CCCTGGAGCA GATGCGCAAG TTCCTTTCGG ACCCGCACGT CCACACGGTG CTGGTGGAGC 360  
GCTCCACGCT CAAAGTGGAC GTCGGTGATG AAGGAGAAGA AGAAAAAGAA TTCATTTCCT 420  
ATAACATCAA CNTAGACATT CACTATGGGG TTAAATCCAA TAGCTTGGCA TTCATTAAAC 480  
GTACTCCCGT GATTGATGCA GATAAACCCG TGTCTTCTCA NCTCCGGGTC CTTACACTCA 540  
GTGAANACTC NCCCTACNAA AACTTTGCAT TCTTTCATTA ACAATGCAGT GGCTCCTTTT 600  
TTTAANTCCT ACATTAAAAA ATCTGGCAAG GCAAACAGGG ATGGTGATAA AATGGCTCCT 660  
TCCNTTGAAA AAAAAATTGC CGAACTCNAA ATNGGACTCC TTCCCTTGCA NCAAAATTTT 720  
TGAAATTCCG GAAATCANC CTGCCCAATT CCTCCCC 757

&lt;210&gt; 16

&lt;211&gt; 300

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 16

ATCATTTTCCT TATTTATATT TCATGTTGGA ATGCTTAAAT CGATAACCTT TGTATTTTGA 60  
AGTGCGCGAC ATGGAAGGTG ATCTGCAAGA GCTGCATCAG TCAAACACCG GGGGATAAAT 120  
CTGGATTTGG GTTCCGGCGT CAAGGTGAAG ATAATACCTA AAGAGGAACA CTGTAAAATG 180  
CCAGAAGCAG GTGAANAGCA ACCACAAGTT TAAATGAAGA CAAGCTGAAA CAACGCAAGC 240  
TGGTTTTATA TTAGATATTT GACTTAAACT ATCTCAATAA AGTTTTGCAG CTTTCACCAC 300

&lt;210&gt; 17

13/14

&lt;211&gt; 313

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 17

AAAGATGGCG GCGGGGGAGG TAGGCAGAGC AGGACGCCGC TGCTGCCGCC GCCACCGCCG 60  
CCTCCGCTCC AGTCGCCTCC GGTCCTTCAA ACTCACACCT CCCGGGAGGA GCTGTCCTGG 120  
CGCCGGGTCC CGCGGGGAAA ATGGTGGAGC CAGGGCAAGA TTTACTGCTT GCTGCTTTGA 180  
GTGAGAGTGG AATTAGTCCG AATGACTCTT TGATATTGAT GGTGGAGATG CANGGCTTGC 240  
AACTCCAATG CCTACCCCGT CAGTTCAGCA NTCAGTGCCA CTTANTGCAT TANAAC TANG 300  
TTTGGAGACC GAA 313

&lt;210&gt; 18

&lt;211&gt; 667

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 18

ACTGCCGGGC TCGGCGTGAG TCGCTGCGGG GCTGACGGGG TGGCAGTGCG GCGGGTTACG 60  
GCCTGGTCAG ACCATAATGA CTTCAGCAAA TAAAGCAATC GAATTACAAC TACAAGTGAA 120  
ACAAAATGCA GAAGAATTAC AAGACTTTAT GCGGGATTTA GAAAACTGGG AAAAAGACAT 180  
TAAACAAAAG GATATGGAAC TAAGAAGACA GAATGGTGTT CCTGAAGAGA ATTTACCTCC 240  
TATTCGAAAT GGGAATTTTA GGAAAAAGAA GAAAGGCAAA GCTAAAGAGT CTTCCCCAAA 300  
ACCANAGAGG AAAACACNAA AAACAGGATA AAATCTTATG ATTATGANGC ATGGGCAAAA 360  
CTTGATGTGG ACCGTATCCT TGATGAGCTT GACAAAGACG ATAGTACCCA TGAGTCTCTG 420

14/14

TCTCAAGAAT CAGAGTCGGA AGAAGATGGG ATTCATGTTG ATTCNCNAAA GGCTCTTGTT 480  
TTAAAAGAAA AGGGCNATAA ATACTTCCAC AAGGAAAATA TGATGAAGCA ATTGACTGCT 540  
ACACNAAAGG CNTGGATGCC GATCCATATN ATCCCGTGTT GCCAACGAAC ANAACNTCCG 600  
CATATTTTAG ACTGAAAAAA TTTGCTGTTG CTGAATCTGA TTGTTATTAN CANTTGCCT 660  
TGAAATA 667

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> Int.Cl <sup>6</sup> C12N15/10		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b>		
Minimum documentation searched (classification system followed by classification symbols) Int.Cl <sup>6</sup> C12N15/10		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) BIOSIS (DIALOG), WPI/L (DIALOG)		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO, 94/08001, A1 (The Kanagawa Academy of Science), 14 April, 1994 (14. 04. 94) & JP, 6-153953, A & EP, 625572, A1	1-7
A	WO, 96/34981, A2 (GENSET), 7 November, 1996 (07. 11. 96) & EP, 824598, A2 & FR, 2733762, A1 & FR, 2733765, A1	1-7
A	Maruyama, K., et al., "Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribo-nucleotides", Gene, Vol. 138 (1994), p.171-174	1-7
A	Kato, S., et al., "Construction of a human full-length cDNA bank", Gene, Vol. 150 (1994), p.243-250	1-7
A	Carnincle, P., et al., "High-Efficiency Full-Length cDNA Cloning by Biotinylated CAP Trapper", (Genomics, Vol. 37 (1996), p.327-336	1-7
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family	
Date of the actual completion of the international search 13 January, 1999 (13. 01. 99)	Date of mailing of the international search report 26 January, 1999 (26. 01. 99)	
Name and mailing address of the ISA/ Japanese Patent Office	Authorized officer	
Facsimile No.	Telephone No.	

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Ederly, I., et al., "An Efficient Strategy To Isolate Full-Length cDNAs Based on an mRNA Cap Retention Procedure (CAPture)", Molecular and Cellular Biology, Vol. 15 (1995), p.3363-3371	1-7
A	Solovyev, V., et al., "Predicting internal exons by oligo-nucleotide composition and discriminant analysis of spliceable open reading frames", Nucleic Acids Research, Vol. 22 (1994), No. 24, p.5156-5163	1-7
A	Heindell, H.C., et al., "The Primary Sequence of Rabbit $\alpha$ -Globin mRNA", Cell, Vol. 15 (1978), p.43-54	1-7
A	Minoru Suzuki et al., "RT-PCR Process: Cloning of 5' end of mRNA by Oligocapping Procedure (in Japanese)", Protein, Nucleic Acid and Enzyme, Vol. 41, No. 5 (1996), p.603-607	1-7
A	Sumio Sugano et al., "Aiming at Full-length cDNA Library: Substitution of Capped Structure by Oligonucleotide (in Japanese)", Protein, Nucleic Acid and Enzyme, Vol. 38, No. 3 (1993), p.476-481	1-7
A	Carninci, P., et al., "High Efficiency Selection of Full-length cDNA by Improved Biotinylated Cap Trapper", DNA Research, Vol. 4, No. 1 (1997), p.61-66	1-7

A. 発明の属する分野の分類 (国際特許分類 (IPC)).  
Int. Cl<sup>o</sup> C12N15/10

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))  
Int. Cl<sup>o</sup> C12N15/10

最小限資料以外の資料で調査を行った分野に含まれるもの

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)  
BIOSIS (DIALOG)、WPI/L (DIALOG)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	WO, 94/08001, A1 (財団法人神奈川科学技術アカデミ ー), 14. 4月. 1994 (14. 04. 94) & JP, 6-1 53953, A&EP, 625572, A1	1-7
A	WO, 96/34981, A2 (GENSET), 7. 11月. 1 996 (07. 11. 96) & EP, 824598, A2&FR, 2733762, A1&FR, 2733765, A1	1-7
A	Maruyama, K., et al. "Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribo- nucleotides", Gene, Vol. 138(1994), p. 171-174	1-7

☒ C欄の続きにも文献が列挙されている。

☐ パテントファミリーに関する別紙を参照。

\* 引用文献のカテゴリー

「A」 特に関連のある文献ではなく、一般的技術水準を示すもの

「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの

「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)

「O」 口頭による開示、使用、展示等に言及する文献

「P」 国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの

「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

「&」 同一パテントファミリー文献

国際調査を完了した日

13. 01. 99

国際調査報告の発送日

26.01.99

国際調査機関の名称及びあて先

日本国特許庁 (ISA/J P)

郵便番号100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

村上 騎見高

印

4 B

8827

電話番号 03-3581-1101 内線 3448

様式PCT/ISA/210 (第2ページの続き) (1998年7月)

---

**THIS PAGE BLANK (USPTO)**